6/12/03

REMARKS

I. INTRODUCTION

In response to the Office Action dated April 16, 2003, claims 2, 10 and 18 have been cancelled, and claims 1, 9 and 17 have been amended. Claims 1, 3-9, 11-17 and 19-24 remain in the application. Entry of these amendments, and re-consideration of the application, as amended, is requested.

II. PRIOR ART REJECTIONS

A. The Office Action Rejections

In paragraphs (1)-(2) of the Office Action, claims 1, 3, 7, 9, 11, 15, 17, 19, and 23 were rejected under 35 U.S.C. §102(e) as being anticipated by Fayyad et al., U.S. Patent No. 6,263,337 (Fayyad). In paragraphs (3)-(4) of the Office Action, claims 2, 4-6, 10, 12-14, 18, and 20-22 were rejected under 35 U.S.C. §103(a) as being unpatentable over Fayyad in view of Van Huben et al., U.S. Patent No. 6,327,594 (Van Huben). In paragraph (5) of the Office Action, claims 8, 16, and 24 were rejected under 35 U.S.C. §103(a) as being unpatentable over Fayyad in view of Guha et al., U.S. Patent No. 6,049,797 (Guha).

Applicants' attorney respectfully traverses these rejections.

B. The Applicants' Independent Claims

Independent claim 1 is directed to a data structure for analyzing data in a computer-implemented data mining system, wherein the data structure is a data model that comprises a Gaussian Mixture Model that stores transactional data, a basket table that contains summary information about the transactional data, an item table that contains information about individual items referenced in the transactional data, and a department table that contains aggregate information about the transactional data, and the data model is mapped to aggregate the transactional data for cluster analysis.

Independent claim 9 is directed to a method for analyzing data in a computer-implemented data mining system, comprising:

generating a data structure in the computer-implemented data mining system, wherein the data structure is a data model that comprises a Gaussian Mixture Model that stores transactional data, a basket table that contains summary information about the transactional data, an item table

that contains information about individual items referenced in the transactional data, and a department table that contains aggregate information about the transactional data; and

mapping the data model to aggregate the transactional data for cluster analysis.

Independent claim 17 is directed to an apparatus for analyzing data in a computer-implemented data mining system, comprising:

means for generating a data structure in the computer-implemented data mining system, wherein the data structure is a data model that comprises a Gaussian Mixture Model that stores transactional data, a basket table that contains summary information about the transactional data, an item table that contains information about individual items referenced in the transactional data, and a department table that contains aggregate information about the transactional data; and

means for mapping the data model to aggregate the transactional data for cluster analysis.

C. The Fayyad Reference

Fayyad describes one exemplary embodiment providing a data mining system for use in finding clusters of data items in a database or any other data storage medium. Before the data evaluation begins a choice is made of the number M of models to be explored, and the number of clusters (K) of clusters within each of the M models. The clusters are used in categorizing the data in the database into K different clusters within each model. An initial set of estimates for a data distribution of each model to be explored is provided. Then a portion of the data in the database is read from a storage medium and brought into a rapid access memory buffer whose size is determined by the user or operating system depending on available memory resources. Data contained in the data buffer is used to update the original model data distributions in each of the K clusters over all M models. Some of the data belonging to a cluster is summarized or compressed and stored as a reduced form of the data representing sufficient statistics of the data. More data is accessed from the database and the models are updated. An updated set of parameters for the clusters is determined from the summarized data (sufficient statistics) and the newly acquired data. Stopping criteria are evaluated to determine if further data should be accessed from the database.

D. The Van Huben Reference

Van Huben describes a common access method to enable disparate pervasive computing devices to interact with centralized data management systems. A modular, scalable data management system is envisioned to further expand the role of the pervasive devices as direct participants in the

data management system. This data management system has a plurality of data managers and is provided with a plurality of data managers in one or more layers of a layered architecture. The system performs with a data manager and with a input from a user or pervasive computing device via an API a plurality of process on data residing in heterogeneous data repositories of computer system including promotion, check-in, check-out, locking, library searching, setting and viewing process results, tracking aggregations, and managing parts, releases and problem fix data under management control of a virtual control repository having one or more physical heterogeneous repositories. The system provides for storing, accessing, tracking data residing in said one or more data repositories managed by the virtual control repository. DMS applications executing directly within, on or behalf of, the pervasive computing device organize data using the PFVL paradigm. Configurable managers include a query control repository for existence of peer managers and provide logic switches to dynamically interact with peers. A control repository layer provides a common process interface across all managers. A command translator performs the appropriate mapping of generic control repository layer calls to the required function for the underlying storage engine.

E. The Guha Reference

Guha describes an invention relating to a computer method, apparatus and programmed medium for clustering databases containing data with categorical attributes. The present invention assigns a pair of points to be neighbors if their similarity exceeds a certain threshold. The similarity value for pairs of points can be based on non-metric information. The present invention determines a total number of links between each cluster and every other cluster bases upon the neighbors of the clusters. A goodness measure between each cluster and every other cluster based upon the total number of links between each cluster and every other cluster and the total number of points within each cluster and every other cluster is then calculated. The present invention merges the two clusters with the best goodness measure. Thus, clustering is performed accurately and efficiently by merging data based on the amount of links between the data to be clustered.

F. The Applicants' Claims Are Patentable Over The References

Applicants' invention, as recited in independent claims 1, 9 and 17, is patentable over the references, because the claims recite limitations not found in the references. Specifically, the combination of Fayyad and Van Huben does not disclose a data model that comprises a Gaussian

Mixture Model that stores transactional data, a basket table that contains summary information about the transactional data, an item table that contains information about individual items referenced in the transactional data, and a department table that contains aggregate information about the transactional data, and the data model is mapped to aggregate the transactional data for cluster analysis.

The Office Action cites Fayyad as teaching all the elements of the independent claims, including a data structure for analyzing data in a computer-implemented data mining system, as reference number 12 in FIG. 2 and in the accompanying text. The Office Action also cites Fayyad as teaching that the data structure is a data model that comprises a Gaussian Mixture Model that stores transactional data, at col. 9, lines 22-67. In addition, the Office Action cites Fayyad as teaching that the data model is mapped to aggregate the transactional data for cluster analysis, at col. 8, lines 34-46. With regard to the method and apparatus claims, the Office Action cites Fayyad as teaching generating the data structure referred to above, at col. 9, line 57 to col. 11, line 29.

Finally, with regard to dependent claims 2, 10 and 18, which are now incorporated into the independent claims, the Office Action cites Fayyad at col. 11, lines 53-67 (for "the data model includes a basket table that contains summary information about the transactional data"), Van Huben at col. 23, lines 7-26 (for "the data model includes ... an item table that contains information about individual items referenced in the transactional data"), and Van Huben at col. 25, lines 49-63 (for "the data model includes ... a department table that contains aggregate information about the transactional data").

Applicants' attorney disagrees. At the locations indicated above, Fayyad and Van Huben, taken individually or in combination, do not teach the claim limitations directed to a data model comprising a Gaussian Mixture Model that stores transactional data, a basket table that contains summary information about the transactional data, an item table that contains information about individual items referenced in the transactional data, and a department table that contains aggregate information about the transactional data, and the data model is mapped to aggregate the transactional data for cluster analysis:

> Fayyad: col. 11, lines 53-67 (actually col. 11, line 53 – col. 12, line 5)
>
> Consider the data points in table 1 again. Assume that the two clusters G1 and G2 of FIG. 5 represent two data clusters after a number of iterations for the age attribute of the table 1 data. After multiple data gathering steps the means of the clusters are 39 and 58 yrs respectively.

data management system. This data management system has a plurality of data managers and is provided with a plurality of data managers in one or more layers of a layered architecture. The system performs with a data manager and with a input from a user or pervasive computing device via an API a plurality of process on data residing in heterogeneous data repositories of computer system including promotion, check-in, check-out, locking, library searching, setting and viewing process results, tracking aggregations, and managing parts, releases and problem fix data under management control of a virtual control repository having one or more physical heterogeneous repositories. The system provides for storing, accessing, tracking data residing in said one or more data repositories managed by the virtual control repository. DMS applications executing directly within, on or behalf of, the pervasive computing device organize data using the PFVL paradigm. Configurable managers include a query control repository for existence of peer managers and provide logic switches to dynamically interact with peers. A control repository layer provides a common process interface across all managers. A command translator performs the appropriate mapping of generic control repository layer calls to the required function for the underlying storage engine.

E. The Guha Reference

Guha describes an invention relating to a computer method, apparatus and programmed medium for clustering databases containing data with categorical attributes. The present invention assigns a pair of points to be neighbors if their similarity exceeds a certain threshold. The similarity value for pairs of points can be based on non-metric information. The present invention determines a total number of links between each cluster and every other cluster bases upon the neighbors of the clusters. A goodness measure between each cluster and every other cluster based upon the total number of links between each cluster and every other cluster and the total number of points within each cluster and every other cluster is then calculated. The present invention merges the two clusters with the best goodness measure. Thus, clustering is performed accurately and efficiently by merging data based on the amount of links between the data to be clustered.

F. The Applicants' Claims Are Patentable Over The References

Applicants' invention, as recited in independent claims 1, 9 and 17, is patentable over the references, because the claims recite limitations not found in the references. Specifically, the combination of Fayyad and Van Huben does not disclose a data model that comprises a Gaussian

Mixture Model that stores transactional data, a basket table that contains summary information about the transactional data, an item table that contains information about individual items referenced in the transactional data, and a department table that contains aggregate information about the transactional data, and the data model is mapped to aggregate the transactional data for cluster analysis.

The Office Action cites Fayyad as teaching all the elements of the independent claims, including a data structure for analyzing data in a computer-implemented data mining system, as reference number 12 in FIG. 2 and in the accompanying text. The Office Action also cites Fayyad as teaching that the data structure is a data model that comprises a Gaussian Mixture Model that stores transactional data, at col. 9, lines 22-67. In addition, the Office Action cites Fayyad as teaching that the data model is mapped to aggregate the transactional data for cluster analysis, at col. 8, lines 34-46. With regard to the method and apparatus claims, the Office Action cites Fayyad as teaching generating the data structure referred to above, at col. 9, line 57 to col. 11, line 29.

Finally, with regard to dependent claims 2, 10 and 18, which are now incorporated into the independent claims, the Office Action cites Fayyad at col. 11, lines 53-67 (for "the data model includes a basket table that contains summary information about the transactional data"), Van Huben at col. 23, lines 7-26 (for "the data model includes ... an item table that contains information about individual items referenced in the transactional data"), and Van Huben at col. 25, lines 49-63 (for "the data model includes ... a department table that contains aggregate information about the transactional data").

Applicants' attorney disagrees. At the locations indicated above, Fayyad and Van Huben, taken individually or in combination, do not teach the claim limitations directed to a data model comprising a Gaussian Mixture Model that stores transactional data, a basket table that contains summary information about the transactional data, an item table that contains information about individual items referenced in the transactional data, and a department table that contains aggregate information about the transactional data, and the data model is mapped to aggregate the transactional data for cluster analysis:

> Fayyad: col. 11, lines 53-67 (actually col. 11, line 53 – col. 12, line 5)
> Consider the data points in table 1 again. Assume that the two clusters G1 and G2 of FIG. 5 represent two data clusters after a number of iterations for the age attribute of the table 1 data. After multiple data gathering steps the means of the clusters are 39 and 58 yrs respectively.

To free up space for a next iteration of data gathering from the database, some of the data in the structure RS is summarized and stored in one of the two data structures CS or DS. (FIGS. 6A, 6B) To define which data points can be safely summarized or compressed, the invention sets up a Bonferroni confidence interval (CI) which defines a multidimensional "box" whose center is the current mean for the K Gaussians defined in the MODEL (FIG. 6D). In one dimension this confidence interval is a span of data both above and below a cluster mean. The confidence interval can be interpreted in the following way: one is confident, to a given level, that the mean of a Gaussian will not lie outside of the CI if it was re-calculated over a different sample of data. A detailed discussion of the determination of the Bonferroni confidence interval is found in Appendix A of this application.

Van Huben: col. 23, lines 7-26 (actually col. 23, lines 3-26)

As the QA inspectors pass or reject the raw materials, they can enter their results into palmtops, laptops or similar devices. Goods that fail QA can be promoted into a REJECT level where several actions may be triggered via Library Processing. For example, the carrier (i.e. UPS) could be automatically notified (202) to come and pick up the defective parts for return to the supplier. Additionally, a fax or e-mail could be automatically sent to the supplier informing them of the pending parts return and requesting replacement parts or credit. The parts that pass QA can are promoted into the INVENTORY level where an UPDATE QUANTITY (203) Library Process can update the Control Repository (18) with the new quantity of parts now available in INVENTORY. Further benefits are derived from the PFVL paradigm because the INVENTORY level can also contain additional information beyond that pertinent to the parts just received. For instance general information such as design specifications, data sheets, price & sales data, drawings, images, etc. can all coexist at this level. These objects may exist as HyperText Markup Language (HTML) files, Portable Document Format (.pdf) files, graphics files in JPEG, AutoCad, Bitmap or other forms, or information stored as fields in a database. Regardless of its location, application of the PFVL paradigm permits uniform access to all of the data.

Van Huben: col. 25, lines 49-63

FIG. 12C depicts how the Package, FileType, Variance, Level, and Version attributes are mapped into the Lotus Notes environment using databases, documents and document fields. Although the aforementioned example uses Lotus Notes to illustrate the principles contained herein, one skilled in the art can appreciate how these concepts along with other concepts such as Configuration Management, Part Number and Release Management, Fix Tracking, and Library Processing can be further implemented using Lotus Notes or any other remote computing or groupware product employing a means of embodying a program of instructions. One can also appreciate how these concepts can be realized in simpler applications such as spreadsheets which provide the capability to enter data tabular format and perform sort and search operations on the fields.

The above portion of Fayyad merely describe that some of the data in the structure RS is summarized and stored in one of the two data structures CS or DS, while the above portions of Van

6/12/03

Huben merely describe that an inventory "level" can also contain additional information beyond that pertinent to the parts just received, such as design specifications, data sheets, price & sales data, drawings, images, etc., and that Package, FileType, Variance, Level, and Version attributes are mapped into the Lotus Notes environment using databases, documents and document fields. However, the Fayyad and Van Huben references, taken individually or in combination, do not describe a data model comprising a Gaussian Mixture Model that stores transactional data, a basket table that contains summary information about the transactional data, an item table that contains information about individual items referenced in the transactional data, and a department table that contains aggregate information about the transactional data, and the data model is mapped to aggregate the transactional data for cluster analysis.

Moreover, Guha fail to overcome these limitations of Fayyad. Recall that Guha was only cited for teaching one row per transaction, and then only against other dependent claims.

Thus, the references do not teach or suggest Applicants' invention. Moreover, the various elements of Applicants' claimed invention together provide operational advantages over the references. In addition, Applicants' invention solves problems not recognized by the references.

Thus, Applicants' attorney submits that independent claims 1, 9 and 17 are allowable over the references. Further, dependent claims 2-8, 10-16 and 18-24 are submitted to be allowable over the references in the same manner, because they are dependent on independent claims 1, 9 and 17, respectively, and thus contain all the limitations of the independent claims. In addition, dependent claims 2-8, 10-16 and 18-24 recite additional novel elements not shown by the references.

III. CONCLUSION

In view of the above, it is submitted that this application is now in good order for allowance and such allowance is respectfully solicited.

Should the Examiner believe minor matters still remain that can be resolved in a telephone interview, the Examiner is urged to call Applicants' undersigned attorney.

Respectfully submitted,

GATES & COOPER LLP
Attorneys for Applicants

Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, California 90045
(310) 641-8797

Date: June 12, 2003

By: _____
Name: George H. Gates
Reg. No.: 33,500

GHG/io